

December 2008 [ CPC13697 ]

## ***The codec world widens***

***Emerging wideband codecs will bring on the next wave of voice quality improvements.***



**By Jean-Marc Valin**  
*Octasic*

Jean-Marc looks at what's on the horizon for VoIP as the market shifts from narrowband to wideband codecs. Standards-based platforms such as CompactPCI and AdvancedTCA have enabled rapid deployment of telecommunications equipment, while encouraging innovation at the application layer. PCI Telecom Mezzanine Cards (PTMCs) and Advanced Mezzanine Cards have been very successful at attacking the carrier infrastructure VoIP market because of their focus on high availability and interoperability.

The spotlight has now shifted. The emphasis is no longer simply on the hardware, but on the software as well. Customers require a solution that they can quickly integrate and deploy to market. The equipment vendor with the most complete and future-proof feature set wins; thus media gateway DSP selection is a key design decision.

The VoIP market has reached maturity in some respects, and carriers are looking for new value-added features that they can use to differentiate their products. They know that the most important aspect of a VoIP product, from an end user's perspective, is voice quality. This includes high-quality echo cancellation, low packet processing latency, and high-quality speech codecs. The most visible market trend is the shift from narrowband codecs to wideband codecs. It is therefore imperative that system designers be aware of the technology landscape in order to address customer demands.

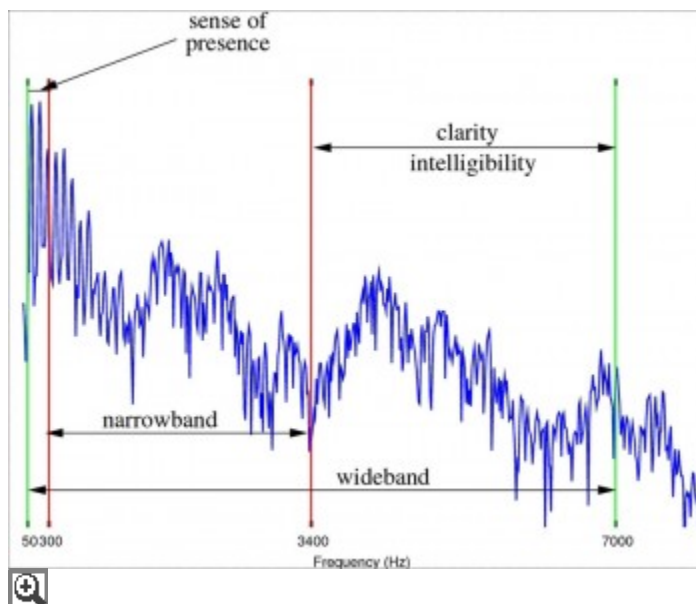
### **Narrowband versus wideband**

Speech carried over the PSTN is currently sampled at 8 kHz. This means that the speech bandwidth is limited to 4 kHz. In practice, only frequencies between 300 Hz and 3.4 Hz are transmitted. While mostly intelligible, this narrowband transmission offers rather poor quality. The first generation of VoIP applications employed narrowband technology, inheriting the same quality limitations as the PSTN. What is worse, the extra delay introduced by the codecs and IP transmission meant that the overall quality of service could only be inferior to the PSTN.

Recently, wideband transmission has been gaining popularity in voice transmission applications, such as VoIP and mobile-to-mobile communications that do not use the PSTN. Wideband speech is sampled at 16 kHz and usually contains frequencies from 50 Hz to 7 kHz, resulting in much higher perceived quality. For the first time, VoIP can surpass the overall quality achievable with the PSTN. As depicted in Figure 1, the extra bandwidth at higher frequencies increases the impression of "clarity" and makes it easier to distinguish between sounds such

as "s" and "f". Wideband and narrowband also differ in the low-frequency range. The extra bandwidth from 50-300 Hz adds a sense of "presence" to the voice.

Figure 1 (click image to zoom by 2.5x)



**Figure 1**  
(click image to zoom by 2.5x)

Two main types of audio codecs exist: speech codecs and music codecs.

Speech codecs:

- Designed mainly for real-time speech communications
- Lower delay
- Lower quality
- Typically 8 kHz or 16 kHz

Music codecs:

- Designed mostly for entertainment applications
- Higher delay
- Higher quality (up to "CD-quality")
- Typically 44.1-48 kHz

### **Embedded codecs**

An embedded codec uses layering to provide incrementally better performance while maintaining interoperability. This allows an encoder and decoder to interoperate effectively without the need for transcoding.

For example, a simpler decoder would only have to interpret the relevant layers in order to reconstruct a narrowband signal, ignoring the wideband extensions.

### **G.711.1**

The new G.711.1 codec is an embedded wideband codec built on top of the narrowband G.711 codec. Its aim is to easily interoperate with the legacy G.711 infrastructure. It provides two extension layers. The first extension improves the quality of G.711 in the low frequencies, including the 50-300 Hz band, which is usually not transmitted with G.711. The second extension encodes high frequencies up to 7 kHz using distinct extension layers.

The G.711.1 codec has a low delay of only 5 ms and a low complexity of 8.7 WMOPS. WMOPS, or Weighted Millions of Operations per second, are a generic unit of measure, essentially described as MIPS for a "typical DSP." One of the most compelling features of G.711.1 is that it allows partial mixing of conference calls. Instead of having to decode, mix, and re-encode all streams, a G.711.1 bridge can mix the base G.711 stream and then forward only the extension layers for the active speaker. Of course, this feature comes with a bit-rate cost. While it is now possible to obtain very good wideband quality at 32 Kbps or below with other codecs, wideband G.711 requires 80-96 Kbps. This is usually not an issue for wireline VoIP applications, where bandwidth is not as constrained, but would be unacceptable for wireless applications.

### **G.729.1**

The G.729.1 codec is also an embedded wideband codec, built on top of a G.729 narrowband version. G.729.1 is also "scalable" in that there are many extension layers that are added on top of each other. The first extension layer provides enhanced narrowband quality at 12 Kbps. From there on, extra layers of 2 Kbps each add wideband capability with increasing quality, up to a maximum rate of 32 Kbps. At that rate, G.729.1 is said to provide acceptable music quality. As is the case for G.711.1, the embedded and scalable nature of G.729.1 comes at a cost in quality versus the bit rate needed, and G.729.1 does not perform as well as other wideband codecs, such as G.722.2 (AMR-WB) at equal rate.

### **Open codecs**

As a reaction to the proliferation of patents on audio codecs, a range of open, royalty-free codecs has emerged. This started with the Vorbis codec, a free alternative to MP3. Vorbis was followed by the Speex speech codec that can encode both narrowband and wideband at bit-rates ranging from 4-42 Kbps. Being royalty-free, open codecs help reduce costs and at the same time limit the complications that arise when licensing the final product. On the other hand, several useful technologies in speech coding, such as ACELP, could not be used in Speex due to patent issues.

### **The future of codecs**

Clearly, the trend in the years to come is to continue increasing telecommunications network quality. Some applications have already started moving to super-wideband (32 kHz) or full-bandwidth audio (44.1-48 kHz). At the same time, some applications now require music to be transmitted along with speech. Because of that, the line between speech codecs and music codecs is blurring. Speech codecs such as G.722.1 (Siren) and G.722.2 (AMR-WB) have now evolved into G.722.1C (Siren14) and AMR-WB+, respectively. These codecs, which started on the speech side of the fence, can now transmit music with reasonable quality. On the music side, the Advanced Audio Codec (AAC) has a low-delay profile (AAC-LD) that puts it in direct competition with G.722.1C and AMR-WB+ as a speech codec.

### **Selecting the right DSP solution**

Delivering better voice quality has a direct impact on the bottom line for any equipment manufacturer. Customers often evaluate voice quality as part of their selection process, because they know that otherwise they will be faced with difficult problems when deploying their solutions.

The next wave of voice quality improvements will be brought on by the emerging wideband codecs discussed earlier. When designing a media processing module, it is important to future-proof the product by selecting a DSP solution that supports all the appropriate advanced codecs. Furthermore, selecting a DSP vendor that is involved with standards bodies ensures that the vendor can offer expert advice and help its customers navigate the myriad of emerging codec standards.

*Jean-Marc Valin is a Software Lead Architect at Octasic. He received his PhD in Electrical Engineering from the University of Sherbrooke in 2005. Jean-Marc specializes in speech enhancement, audio coding, and echo cancellation and has published several peer-reviewed papers on these topics. He is also the author of the Speex open source speech codec and a member of the Xiph.Org Foundation and IEEE Signal Processing Society.*

**Octasic**

[jean-marc.valin@octasic.com](mailto:jean-marc.valin@octasic.com)

[www.octasic.com](http://www.octasic.com)